

Искусственный интеллект: из прошлого через будущее в настоящее

Воронцов Константин Вячеславович

д.ф.-м.н., профессор РАН,
зав. кафедрой математических методов прогнозирования МГУ,
рук. лаб. машинного обучения и семантического анализа Института ИИ МГУ,
зав. кафедрой машинного обучения и цифровой гуманитаристики МФТИ,
зав. кафедрой интеллектуальных систем МФТИ,
г.н.с. ФИЦ «Информатика и управление» РАН

k.vorontsov@iai.msu.ru



Докладчик: *Воронцов Константин Вячеславович*

http://www.machinelearning.ru/wiki/index.php?title=User:Vokov

<http://www.MachineLearning.ru/wiki?title=User:Vokov>

Участник:Vokov

Воронцов Константин Вячеславович
профессор РАН, д.ф.-м.н.,
проф., зам. зав. каф. «Математические методы прогнозирования» ВМК МГУ,
зав. каф. «Машинного обучения и цифровой гуманитаристики» МФТИ,
проф. каф. «Интеллектуальные системы» МФТИ,
с.н.с. отдела «Интеллектуальные системы» Вычислительного центра ФИЦ ИУ РАН,
один из идеологов и Администраторов ресурса MachineLearning.RU,
прочие подробности — на подстранице Curriculum vitae.

■ Профиль ORCID = 0000-0002-4244-4270
■ Профиль SCOPUS ID = 6507982932
■ Профиль WoS ResearcherID = G-7857-2014
■ Профиль Google Scholar
■ Профиль DBLP
■ Профиль РИНЦ ID = 15081
■ Профиль в системе ИСТИНА IRID = 3151446
■ Профиль MathNet.ru

Мне можно написать письмо.
<http://www.MachineLearning.ru/wiki?title=User:Vokov>

1 Учебные материалы

Содержание [убрать]

- 1 Учебные материалы
 - 1.1 Курсы лекций
 - 1.2 Рекомендации для студентов и аспирантов
- 2 Интервью
 - 2.1 Видео
 - 2.2 Лонгриды
 - 2.3 Российский радиоуниверситет, Радио России
- 3 Доклады на конференциях и семинарах
- 4 Научные интересы
 - 4.1 Анализ текстов и информационный поиск
 - 4.2 Фейковые новости и потенциально опасный дискурс
 - 4.3 Отслеживание контактов и оценка рисков инфицирования
 - 4.4 Теория обобщающей способности
 - 4.5 Комбинаторная (перестановочная) статистика
 - 4.6 Прогнозирование объемов продаж
 - 4.7 Другие проекты и семинары
- 5 Публикации
- 6 Софт и проекты
- 7 Аспиранты и студенты
 - 7.1 Бакалаврские диссертации
 - 7.2 Магистерские диссертации
 - 7.3 Дипломные работы

Содержание

1. Прошлое

- История развития технологий ИИ
- Генеративный ИИ и большие языковые модели
- Методология и философия ИИ (почему ИИ — это ~~надолго~~ навсегда)

2. Будущее

- Как предсказывать тренды развития технологий?
- От персональных помощников к цифровому посмертию
- Человеко-машинная цивилизация

3. Настоящее

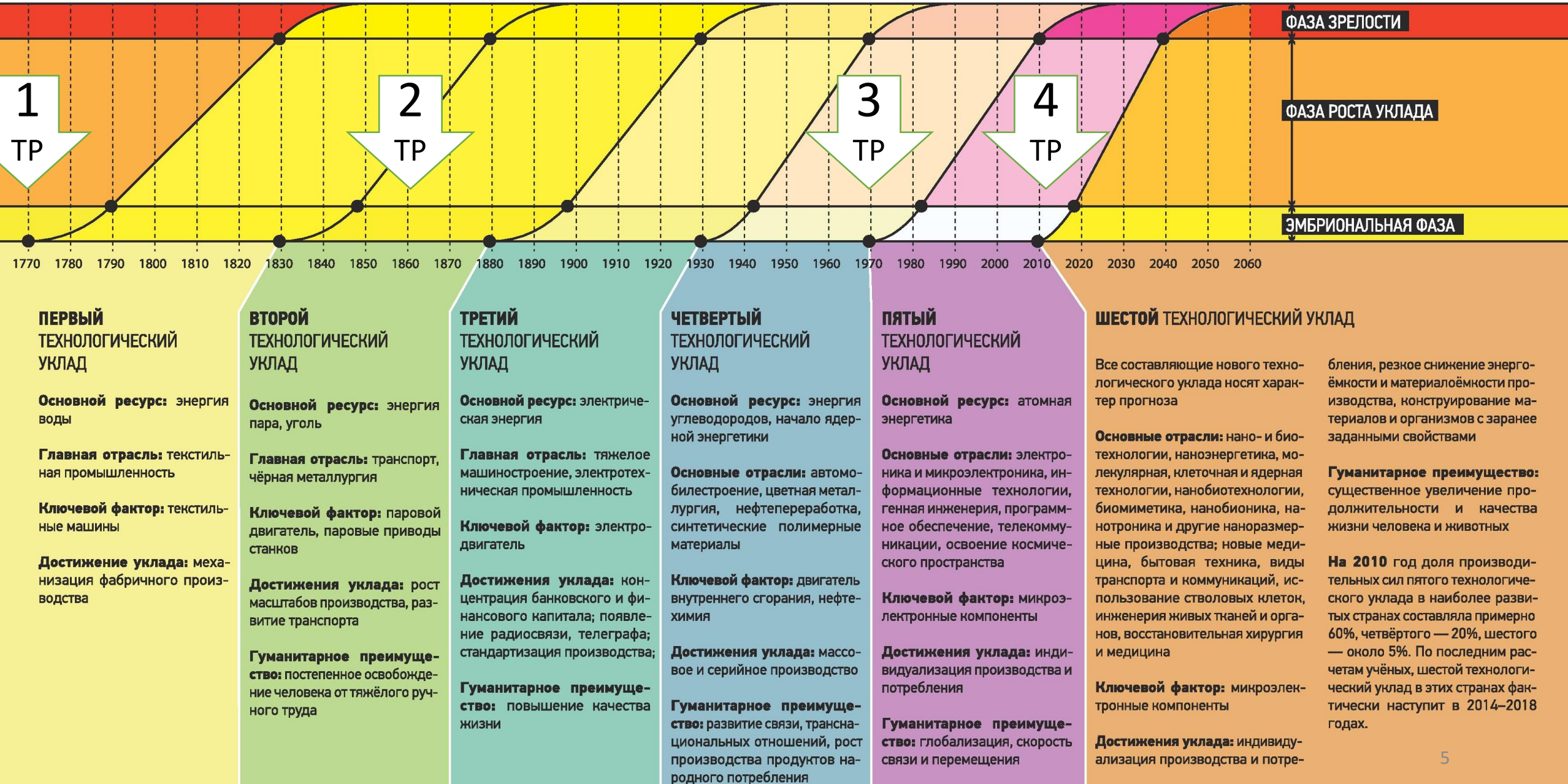
- Целеполагание. Этика. Идеология. Образы будущего
- Передача знаний и человеко-машинная коммуникация
- Открытые данные и конкурсы. Биржи данных. Кадры и образование

«Четвёртая технологическая революция строится на вездесущем и мобильном Интернете, *искусственном интеллекте* и *машинном обучении*» (2016)

Клаус Мартин Шваб,
президент Всемирного
экономического форума



Технологические уклады по Н.Д.Кондратьеву



Отчёты Белого дома США, май-октябрь 2016

«Nations with the strongest presence in AI R&D will establish leading positions in the automation of the future»

Цифровая и распределённая экономика
Автоматизация и сокращение издержек
Автономный транспорт и роботизация
Оптимизация логистики и цепей поставок
Оптимизация энергетических сетей (Energy Tech)
Автоматизация банковских услуг (Fin Tech)
Автоматизация юридических услуг (Legal Tech)
Автоматизация образовательных услуг (Ed Tech)
Автоматизация работы с кадрами (HR Tech)
Персональная медицина (Med Tech)
Автоматизация в сельском хозяйстве (Agro Tech)
Автоматизация в природопользовании (Geo Tech)
Автономные системы вооружений (Mil Tech)



Национальная стратегия развития ИИ в РФ

10 октября 2019 г.



УКАЗ

ПРЕЗИДЕНТА РОССИЙСКОЙ ФЕДЕРАЦИИ

**О развитии искусственного интеллекта
в Российской Федерации**

В целях обеспечения ускоренного развития искусственного интеллекта в Российской Федерации, проведения научных исследований в области искусственного интеллекта, повышения доступности информации и вычислительных ресурсов для пользователей, совершенствования системы подготовки кадров в этой области **п о с т а н о в л я ю:**

1. Утвердить прилагаемую Национальную стратегию развития искусственного интеллекта на период до 2030 года.

2. Правительству Российской Федерации:

а) до 15 декабря 2019 г. обеспечить внесение изменений в национальную программу "Цифровая экономика Российской Федерации", в том числе разработать и утвердить федеральный проект "Искусственный интеллект";

б) представлять Президенту Российской Федерации ежегодно доклад о ходе реализации Национальной стратегии развития искусственного интеллекта на период до 2030 года;

в) предусматривать при формировании в 2020 - 2030 годах проектов федеральных бюджетов на очередной финансовый год и на плановый период бюджетные ассигнования на реализацию настоящего Указа.

УТВЕРЖДЕНА
Указом Президента
Российской Федерации
от 10 октября 2019 г. № 490

**НАЦИОНАЛЬНАЯ СТРАТЕГИЯ
развития искусственного интеллекта
на период до 2030 года**

I. Общие положения

1. Настоящей Стратегией определяются цели и основные задачи развития искусственного интеллекта в Российской Федерации, а также меры, направленные на его использование в целях обеспечения национальных интересов и реализации стратегических национальных приоритетов, в том числе в области научно-технологического развития.

2. Правовую основу настоящей Стратегии составляют Конституция Российской Федерации, Федеральный закон от 28 июня 2014 г. № 172-ФЗ "О стратегическом планировании в Российской Федерации", указы Президента Российской Федерации от 7 мая 2018 г. № 204 "О национальных целях и стратегических задачах развития Российской Федерации на период до 2024 года", от 9 мая 2017 г. № 203 "О Стратегии развития информационного общества в Российской Федерации на 2017 - 2030 годы", от 1 декабря 2016 г. № 642 "О Стратегии научно-технологического развития Российской Федерации" и иные нормативные правовые акты Российской Федерации, определяющие направления применения информационных технологий в Российской Федерации.

3. Настоящая Стратегия является основой для разработки (корректировки) государственных программ Российской Федерации, государственных программ субъектов Российской Федерации, федеральных и региональных проектов, плановых и программно-целевых документов государственных корпораций, государственных компаний, акционерных обществ с государственным участием,



2 100011 98860 6

Бум искусственного интеллекта

1997: IBM Deep Blue обыграл чемпиона мира по шахматам

2005: Беспилотный автомобиль: DARPA Grand Challenge

2006: Google Translate – статистический машинный перевод

2011: 40 лет DARPA CALO привели к созданию Apple Siri

2011: IBM Watson победил в ТВ-игре «Jeopardy!»

2009–2018: ImageNet: 25% → 2,5% ошибок против 5% у людей

2015: Фонд OpenAI в \$1 млрд. Илона Маска и Сэма Альтмана

2016: DeepMind, OpenAI: динамическое обучение играм Atari

2016: Google DeepMind обыграл чемпиона мира по игре го

2017: OpenAI обыграл чемпиона мира по компьютерной игре Dota 2

2020: Модель GPT-3 синтезирует тексты, неотличимые от человеческих

2023: GPT-4 демонстрирует «проблески общего искусственного интеллекта»

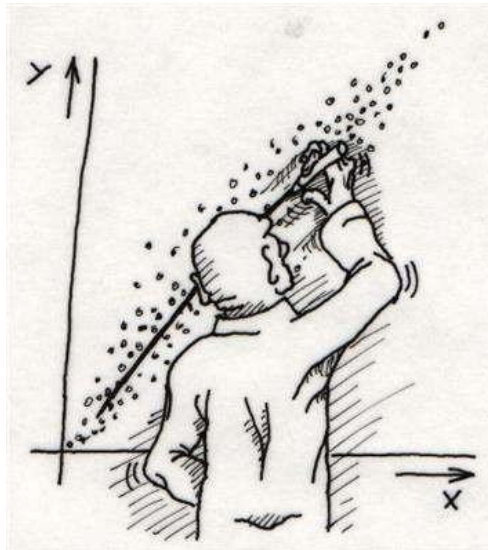
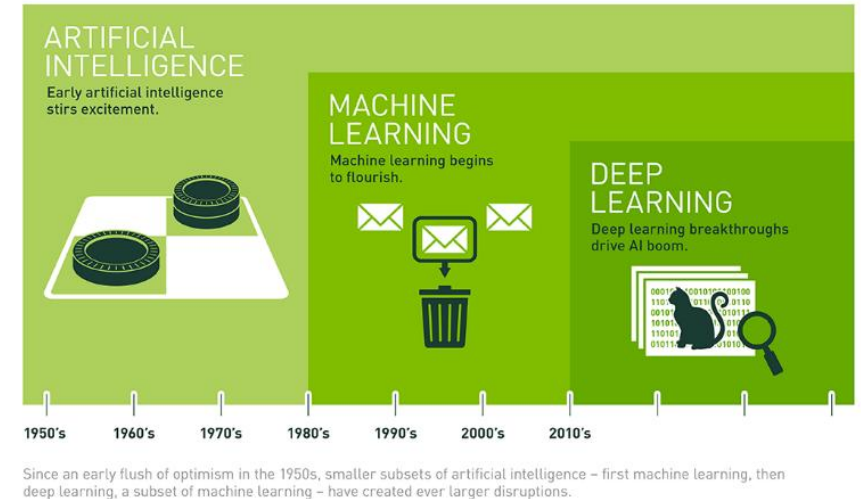


Машинное обучение, большие данные «и много других страшных слов»

- Статистический анализ данных (Statistical Data Analysis)
- Искусственный интеллект (Artificial Intelligence, AI) 1955
- Распознавание образов (Pattern Recognition)
- Машинное обучение (Machine Learning, ML) 1959
- Статистическое обучение (Statistical Learning)
- Интеллектуальный анализ данных (Data Mining) 1989
- Машинный интеллект (Machine Intelligence) 2000
- Бизнес-аналитика (Business Intelligence, Business Analytics)
- Предсказательная аналитика (Predictive Analytics) 2007
- Большие данные (Big Data) 2008
- Аналитика больших данных (Big Data Analytics)
- Наука о данных (Data Science, DS) 2011

Машинное обучение (Machine Learning, ML)

- одна из ключевых информационных технологий будущего
- наиболее успешное направление ИИ, вытеснившее экспертные системы и инженерию знаний



- проведение функции через заданные точки в сложно устроенных пространствах
- математическое моделирование в условиях, когда знаний мало, данных много
- тысячи различных методов и алгоритмов
- более 100 000 научных публикаций в год

Задачи машинного обучения с учителем

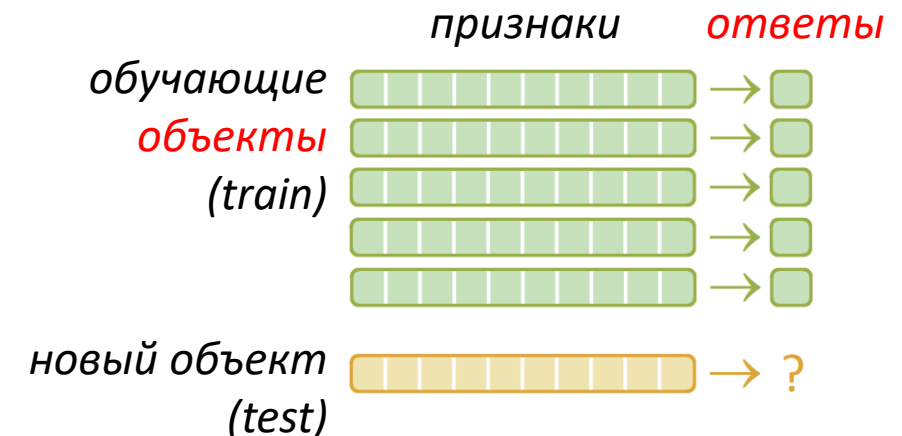
Этап №1 – обучение с учителем

- **На входе:**
данные – выборка прецедентов «**объект** → **ответ**»,
каждый объект описывается набором *признаков*
- **На выходе:**
модель, предсказывающая ответ по объекту

Если нет данных,
то нет
и машинного
обучения

Этап №2 – применение

- **На входе:**
данные – новый **объект**
- **На выходе:**
предсказание **ответа** на новом объекте



Примеры задач ML в бизнесе

- **Кредитный скоринг:**

объект – данные о заёмщике

ответ – решение по кредиту & вероятность дефолта



- **Предсказание оттока клиентов:**

объект – данные о клиенте на момент времени t

ответ – уйдёт ли клиент к моменту времени $t + \Delta$



- **Прогнозирование объёмов продаж:**

объект – данные о продажах на момент времени t

ответ – объём спроса в интервале от t до $t + \Delta$



Примеры задач ML в интернет-сервисах

- **Информационный поиск в Интернете:**

объект – данные о паре «запрос и документ»

ответ – оценка релевантности документа запросу



- **Продажа рекламы в Интернете:**

объект – данные о тройке «пользователь, страница, баннер»

ответ – оценка вероятности клика

- **Рекомендательные системы в Интернете / TV:**

объект – данные о паре «пользователь, товар / фильм»

ответ – оценка вероятности покупки / просмотра



Примеры задач ML

- **Медицинская диагностика:**

объект – данные о пациенте на текущий момент

ответ – диагноз / лечение / риск исхода



- **Поиск месторождений полезных ископаемых:**

объект – данные о геологии района

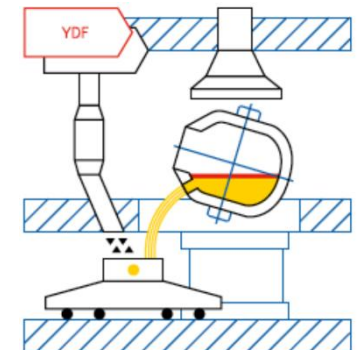
ответ – есть/нет месторождение



- **Управление технологическими процессами:**

объект – данные о сырье и управляющих параметрах

ответ – количество/качество полезного продукта



Примеры задач с не векторными данными

- **Статистический машинный перевод:**

объект – предложение на естественном языке

ответ – его перевод на другой язык

*Прогресс в этих
областях связан с
«**большими данными**»
(англ. «Big Data»)*

- **Перевод речи в текст:**

объект – аудиозапись речи человека

ответ – текстовая запись речи

...очень важное уточнение:

с аккуратными

большими данными

- **Беспилотное вождение:**

объект – поток данных с радаров и видеокамер

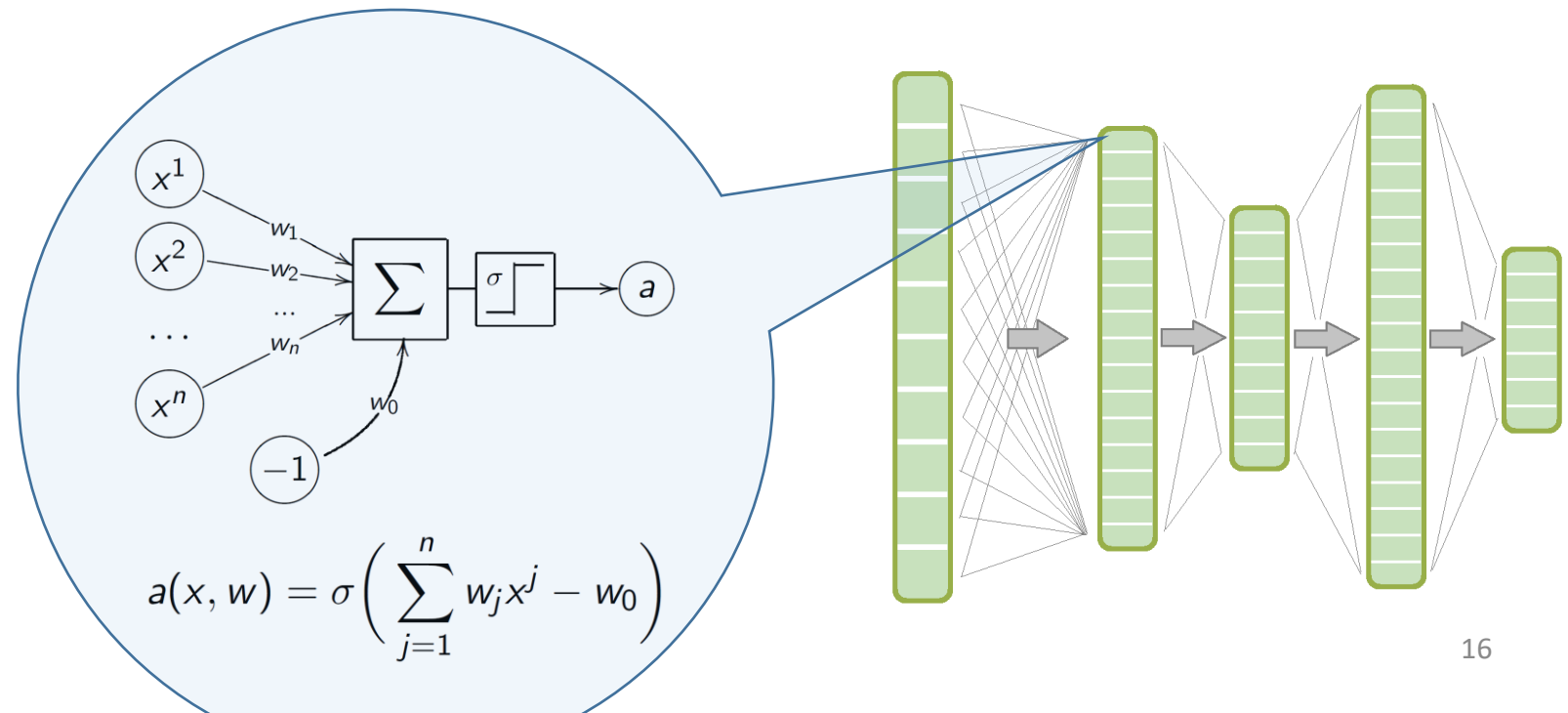
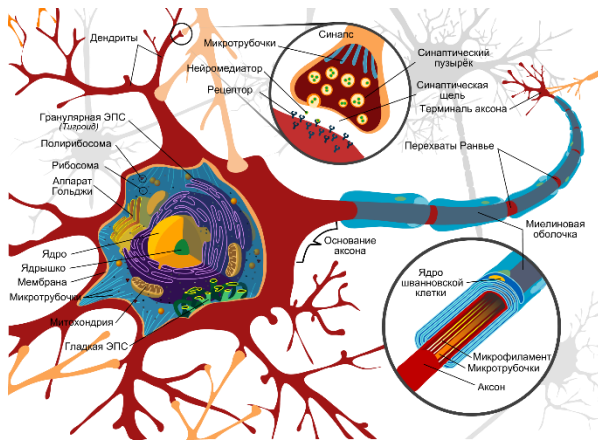
ответ – поток решений (объехать, остановиться, игнорировать, ...)

Модели искусственных нейронных сетей

На каждом слое сети вектор объекта преобразуется в новый вектор

Каждое преобразование (нейрон) – линейная модель $a(x, w)$

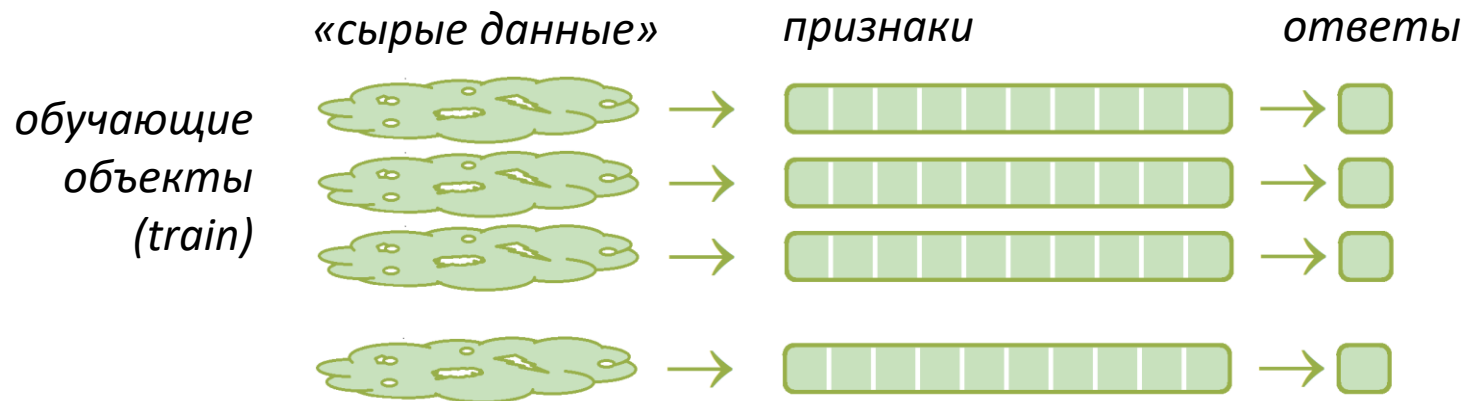
Веса w являются обучаемыми параметрами модели



Глубокие нейронные сети (Deep ANN)

Вход: сложно структурированные «сырые» данные объектов

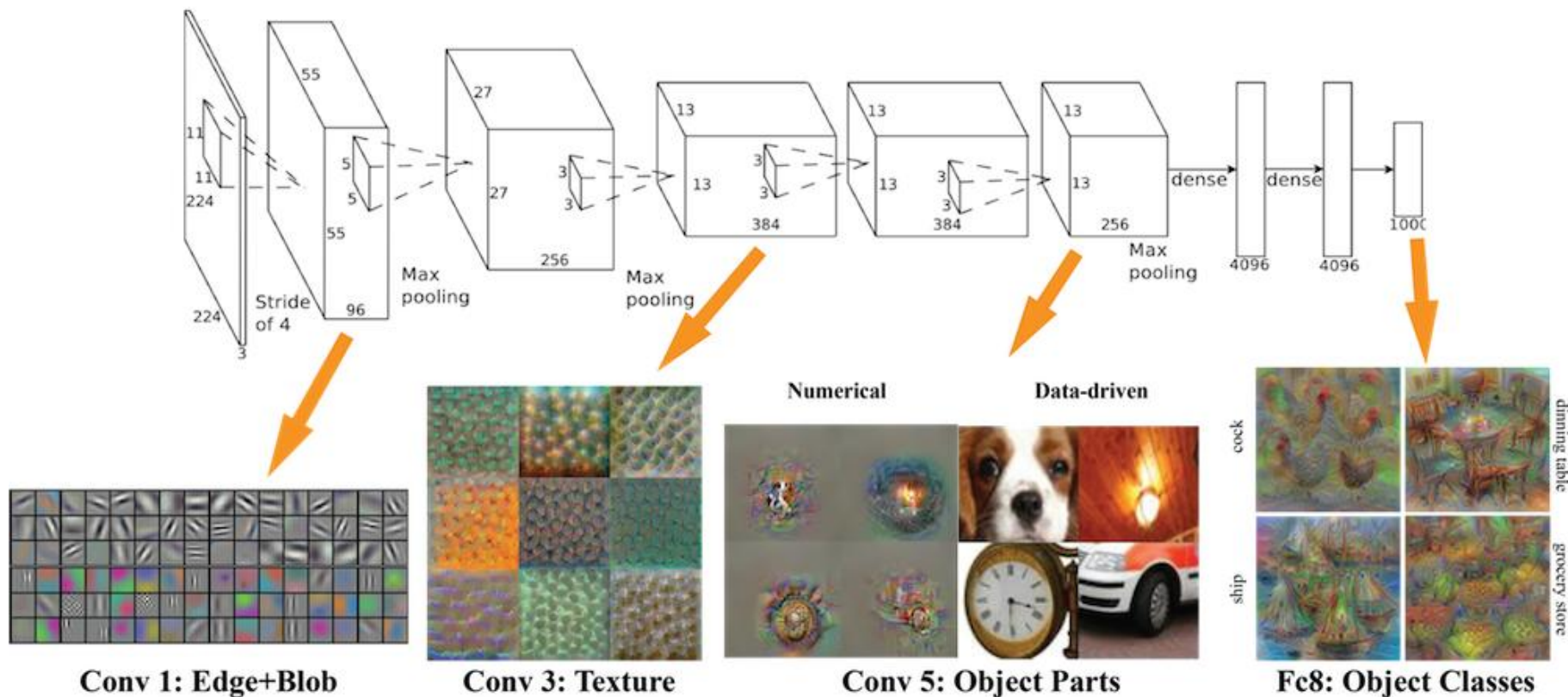
Выход: векторные представления объектов, затем ответы



*Deep Learning – это
всего лишь обучаемая
векторизация
сложных объектов*

Примеры сложно структурированных объектов: изображения, видео, временные ряды, тексты, последовательности, транзакции, графы, ...

Глубокие свёрточные нейронные сети (CNN) для классификации объектов на изображениях

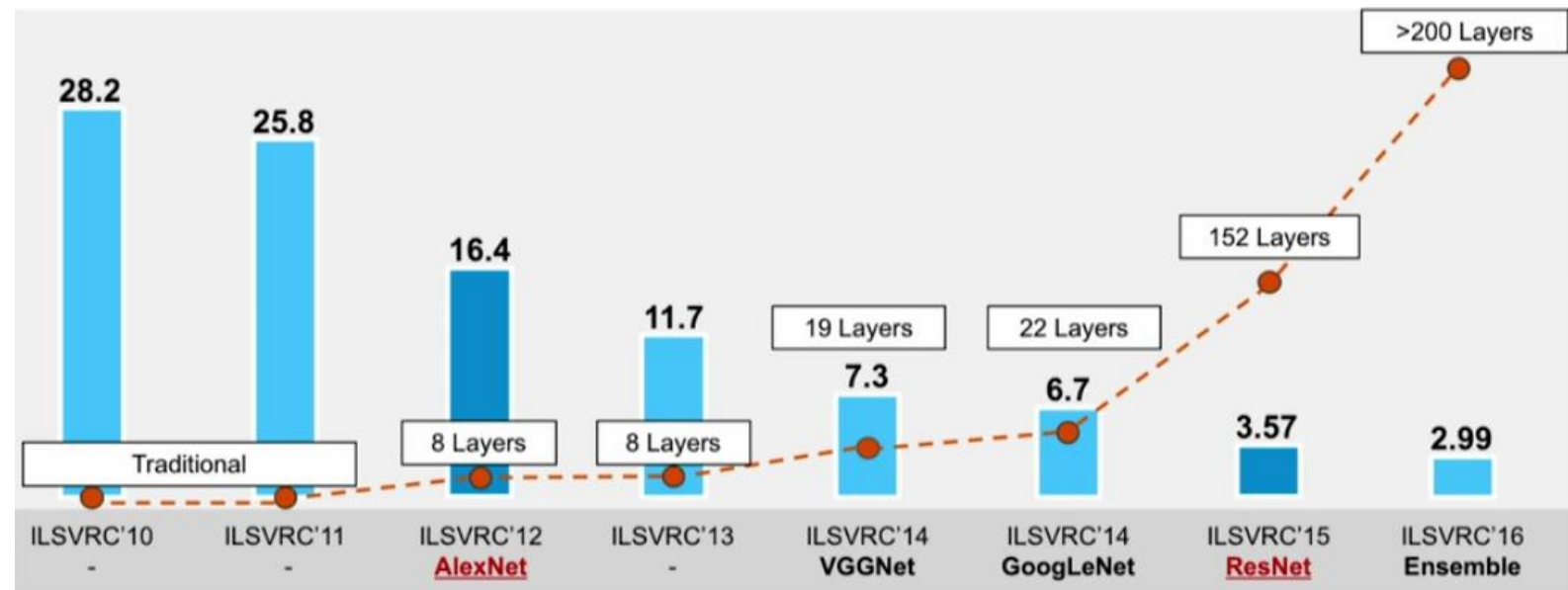


Krizhevsky A., Sutskever I., Hinton G. ImageNet classification with deep convolutional neural networks. 2012.

Роль больших данных

ImageNet: открытая выборка 14М изображений, 20К категорий

IMAGENET



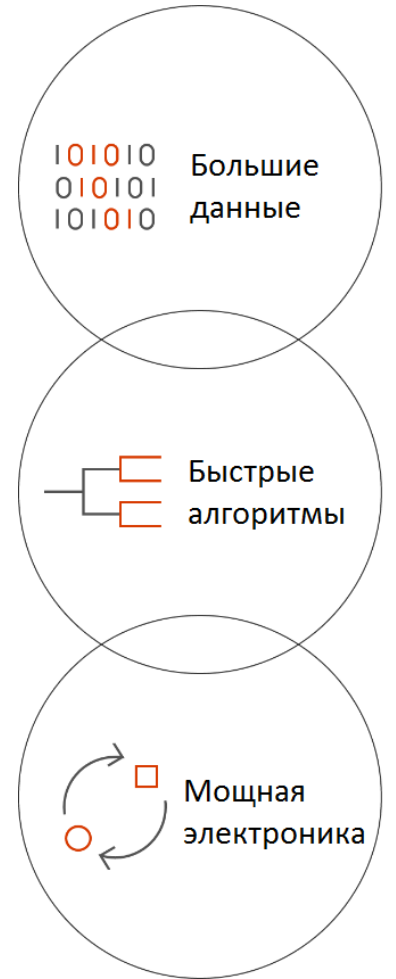
Старт в 2009 г. Человеческий уровень ошибок 5% пройден в 2015 г.

Li Fei-Fei et al. ImageNet: A large-scale hierarchical image database. 2009.

Li Fei-Fei et al. Construction and analysis of a large scale image ontology. 2009.

Три составляющих успеха Deep Learning

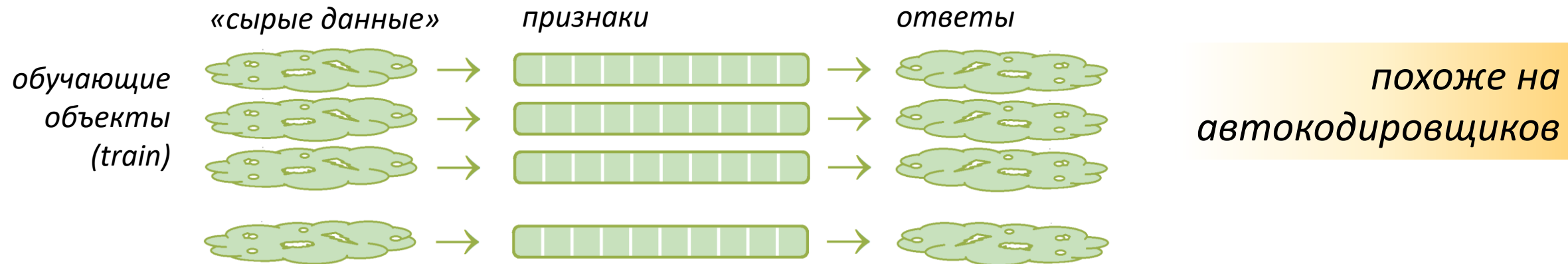
- Повсеместное применение компьютерных технологий
→ *накопление больших выборок данных*
в частности, ImageNet
- Развитие математических методов и алгоритмов
→ *накопление критической массы опыта*
методы оптимизации для больших размерностей
- Достижения микроэлектроники
→ *рост вычислительных мощностей по закону Мура*
в частности, GPU



Нейронные сети для синтеза объектов

Вход: сложно структурированные объекты

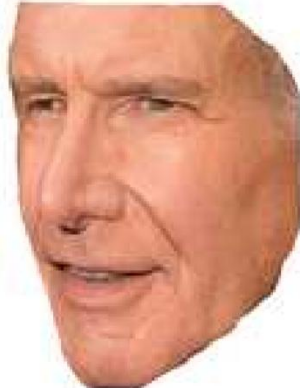
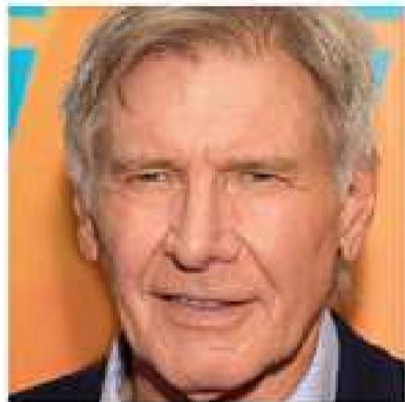
Выход: сложно структурированные ответы



Примеры: синтез изображений, перенос стиля, распознавание речи, машинный перевод, суммаризация текстов, диалог с пользователем

Модели: seq2seq, CNN, RNN, LSTM, GAN, BERT, GPT и др.

Синтез изображений и видео



(d) input image

(e) output 3d face

(f) textured 3d face



Source Subject

Target Subject 1

Target Subject 2

Эволюция подходов в обработке текстов

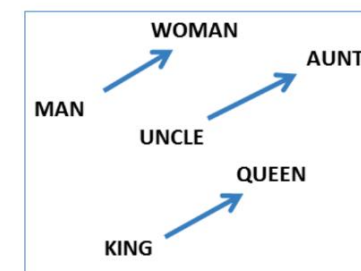
Декомпозиция задач по уровням «пирамиды NLP»

- морфологический анализ, лемматизация, опечатки, ...
- синтаксический анализ, выделение терминов, NER, ...
- семантический анализ, выделение фактов, тем, ...



Модели векторизации слов (эмбедингов)

- модели дистрибутивной семантики: word2vec [Mikolov, 2013], FastText [Bojanowski, 2016], ...
- тематические модели LDA [Blei, 2003], ARTM [2014], ...

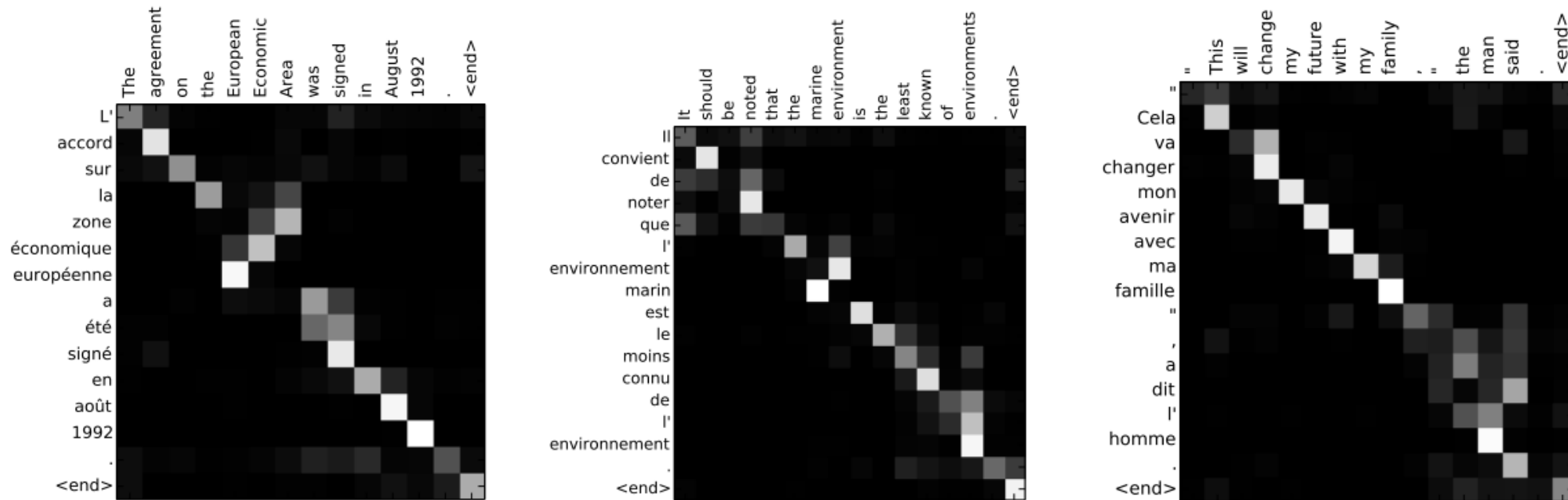


Нейросетевые модели контекстной векторизации

- рекуррентные нейронные сети: LSTM, GRU, ...
- «end-to-end» модели внимания и трансформеры: машинный перевод [2017], BERT [2018], GPT-4 [2023], ...

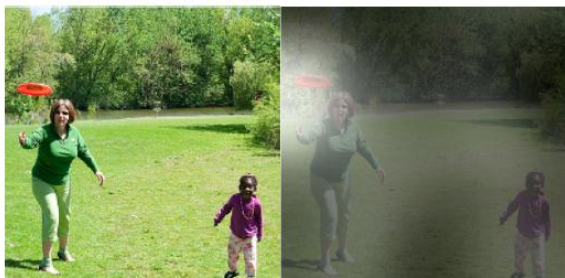
$$\text{softmax} \left(\frac{\begin{matrix} \text{Q} \\ \text{3x3 grid} \end{matrix} \times \begin{matrix} \text{K}^T \\ \text{3x3 grid} \end{matrix}}{\sqrt{d}} \right) \begin{matrix} \text{V} \\ \text{3x1 grid} \end{matrix}$$

Модели внимания: машинный перевод



Интерпретация моделей внимания: *матрица семантического сходства* $A[t,i]$ показывает, на какие слова $x[i]$ входного текста модель обращает внимание, когда генерирует слово перевода $y[t]$

Модели внимания: аннотирование изображений



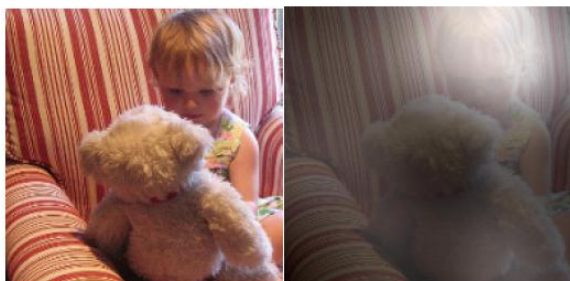
A woman is throwing a frisbee in a park.



A dog is standing on a hardwood floor.



A stop sign is on a road with a mountain in the background.



A little girl sitting on a bed with a teddy bear.



A group of people sitting on a boat in the water.

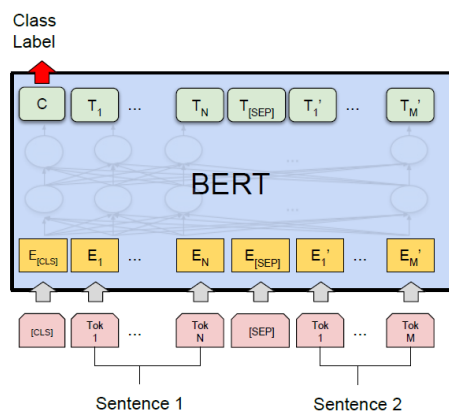


A giraffe standing in a forest with trees in the background.

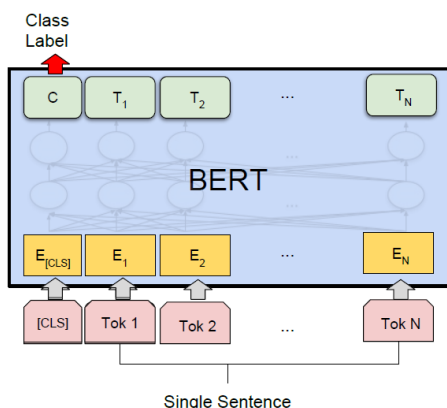
Интерпретация: на какие области модель обращает внимание, генерируя подчёркнутое слово в описании изображения

Трансформеры: большие языковые модели

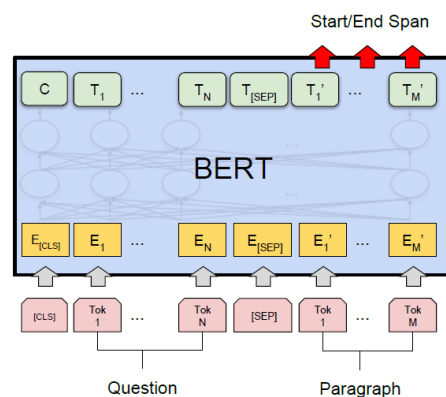
- обучаются векторизовать и предсказывать слова по контексту
- обучаются по терабайтам текстов, «они видели в языке всё»
- мультиязычные: обучаются на десятках языков
- мультизадачные: для каждой новой задачи NLP/NLU достаточно предобученной модели или дообучения на небольшой выборке



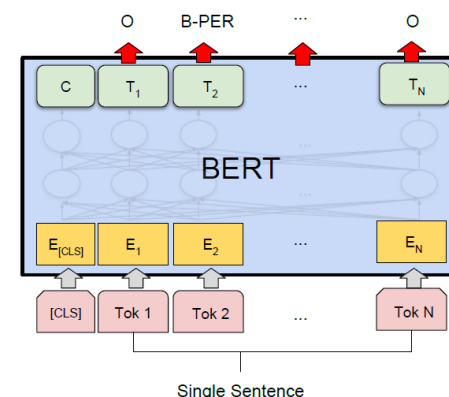
(a) Sentence Pair Classification Tasks:
MNLI, QQP, QNLI, STS-B, MRPC,
RTE, SWAG



(b) Single Sentence Classification Tasks:
SST-2, CoLA



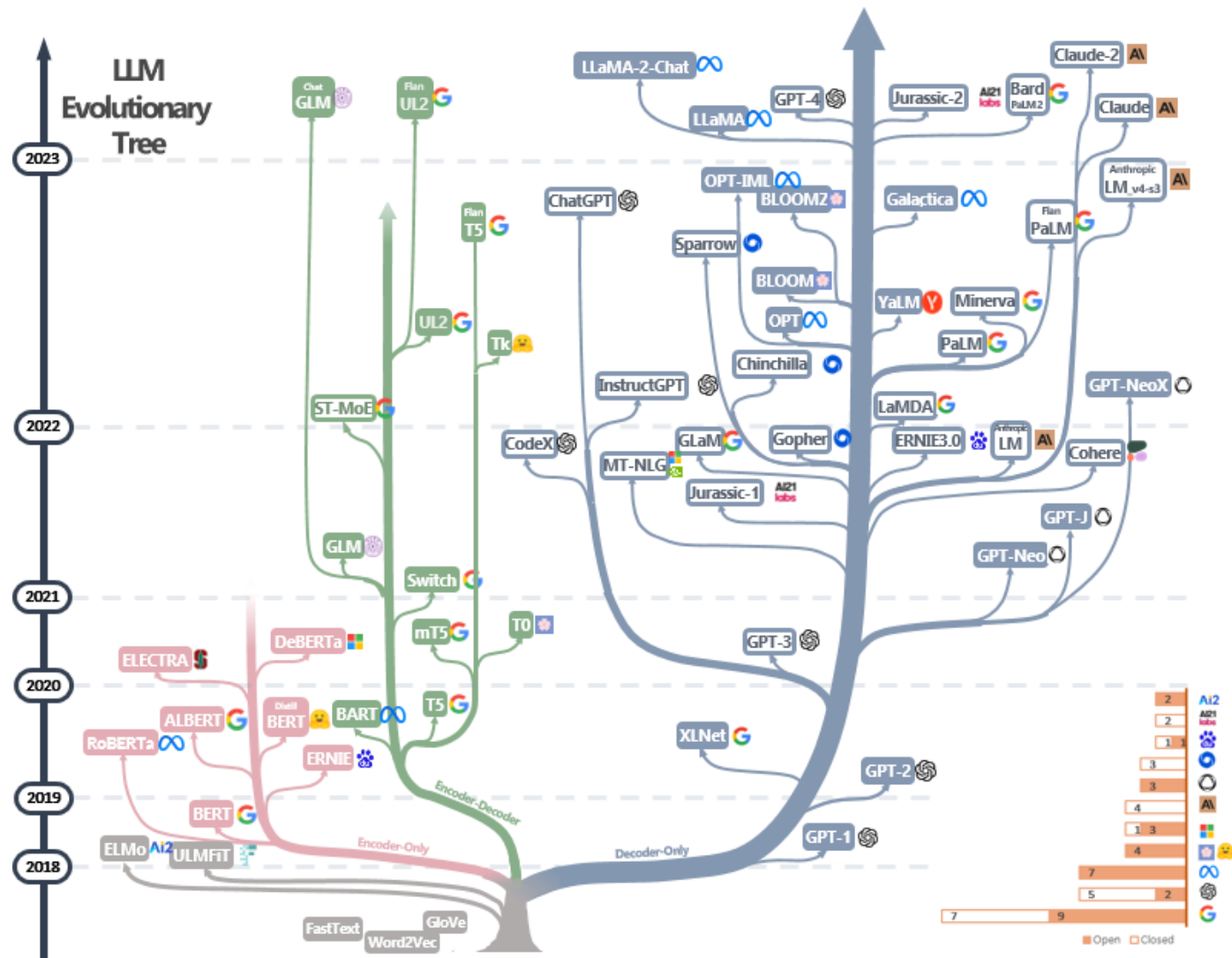
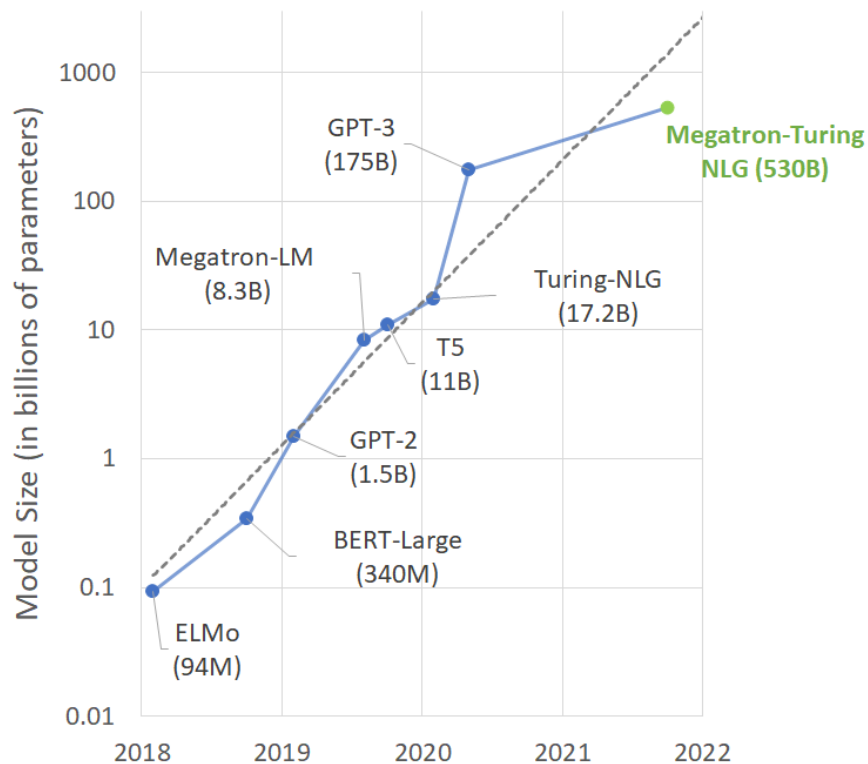
(c) Question Answering Tasks:
SQuAD v1.1



(d) Single Sentence Tagging Tasks:
CoNLL-2003 NER

Трансформеры: размер имеет значение

Рост числа параметров
больших языковых
моделей



Проблески общего искусственного интеллекта

Sparks of Artificial General Intelligence: Early experiments with GPT-4

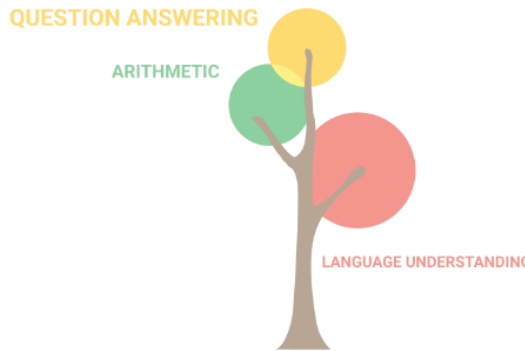
Sébastien Bubeck Varun Chandrasekaran Ronen Eldan Johannes Gehrke
Eric Horvitz Ece Kamar Peter Lee Yin Tat Lee Yuanzhi Li Scott Lundberg
Harsha Nori Hamid Palangi Marco Tulio Ribeiro Yi Zhang

Microsoft Research (27 March 2023)

Новые способности модели GPT, не закладывавшиеся при её обучении:

- объяснять свои ответы, перефразировать, переводить на другие языки
- реферировать, генерировать планы, сценарии, шаблоны
- строить аналогии, менять тональность, стиль, глубину изложения
- генерировать программный код на различных языках
- решать некоторые логические и математические задачи
- искать и исправлять собственные ошибки по подсказке

Новые (эмерджентные) способности модели

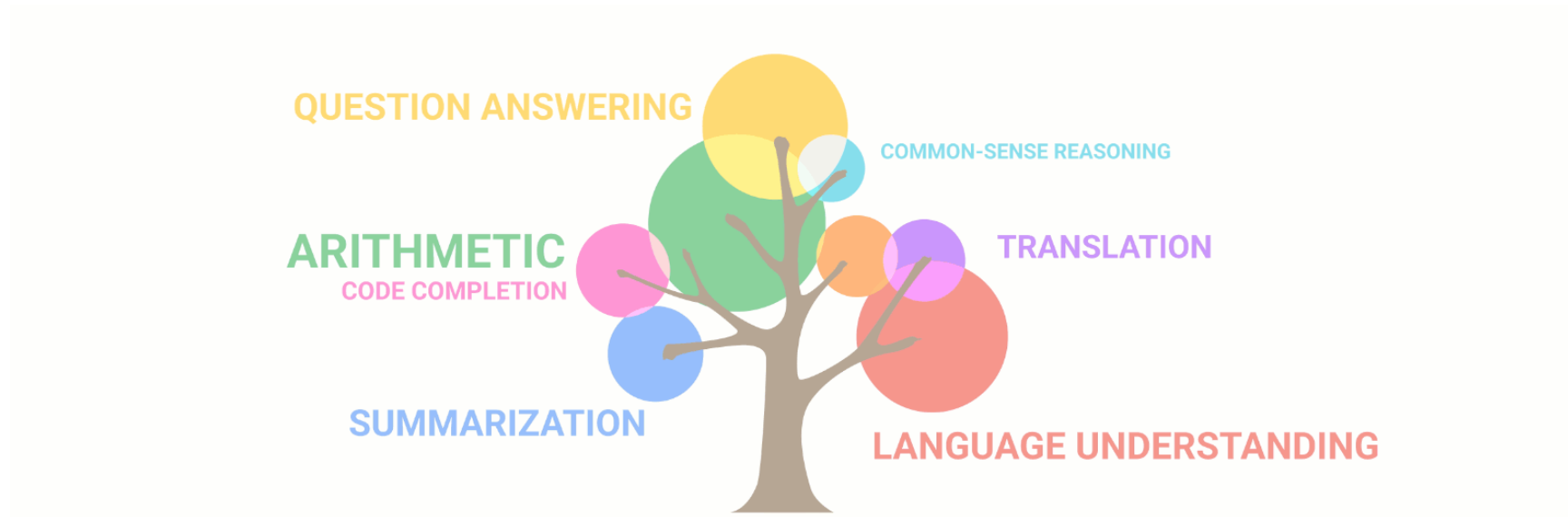


GPT-2: 14-Feb-2019

1,5 млрд. параметров, корпус 10 млрд. токенов (40Gb), контекст 768 слов (1,5 стр.)

- способность написать эссе, которое конкурсное жюри не смогло отличить от написанного человеком

Новые (эмерджентные) способности модели

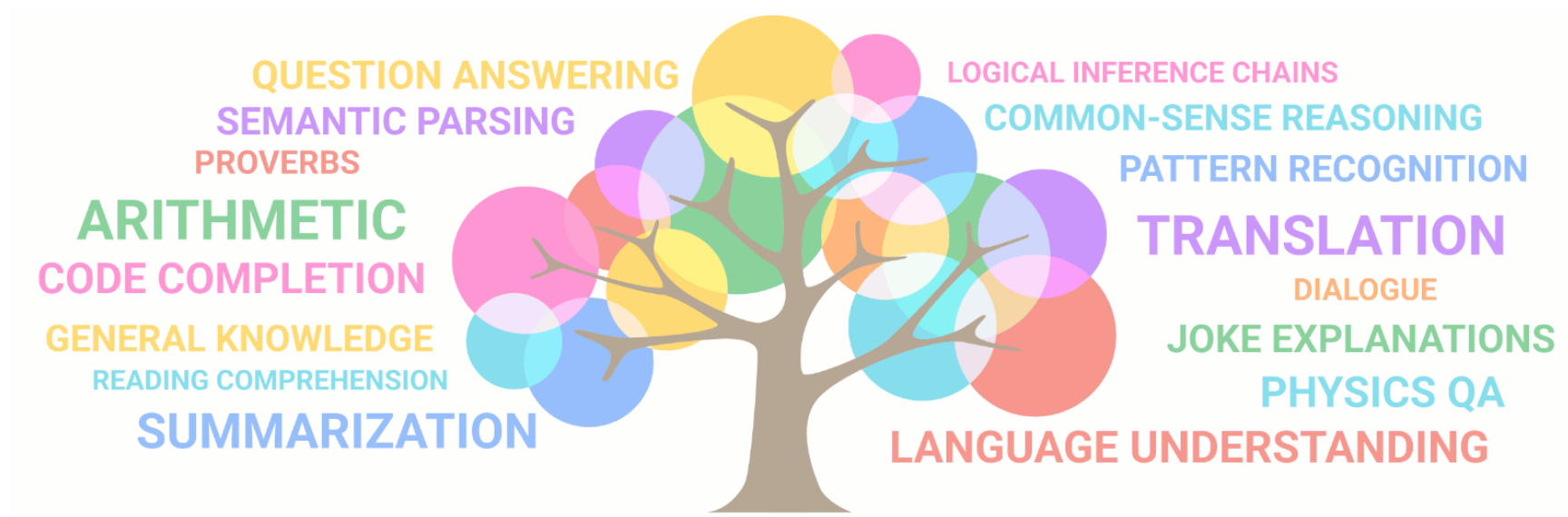


GPT-3: 11-Jun-2020

175 млрд. параметров, корпус 500 млрд. токенов, контекст 1536 слов (3 стр.)

- способность делать перевод на другие языки
- способность решать логические и простейшие математические задачи
- способность генерировать программный код по текстовому описанию

Новые (эмерджентные) способности модели



GPT-4: 14-Mar-2023

>1 трл. параметров, корпус >1Tb, контекст 24 000 слов (48 страниц)

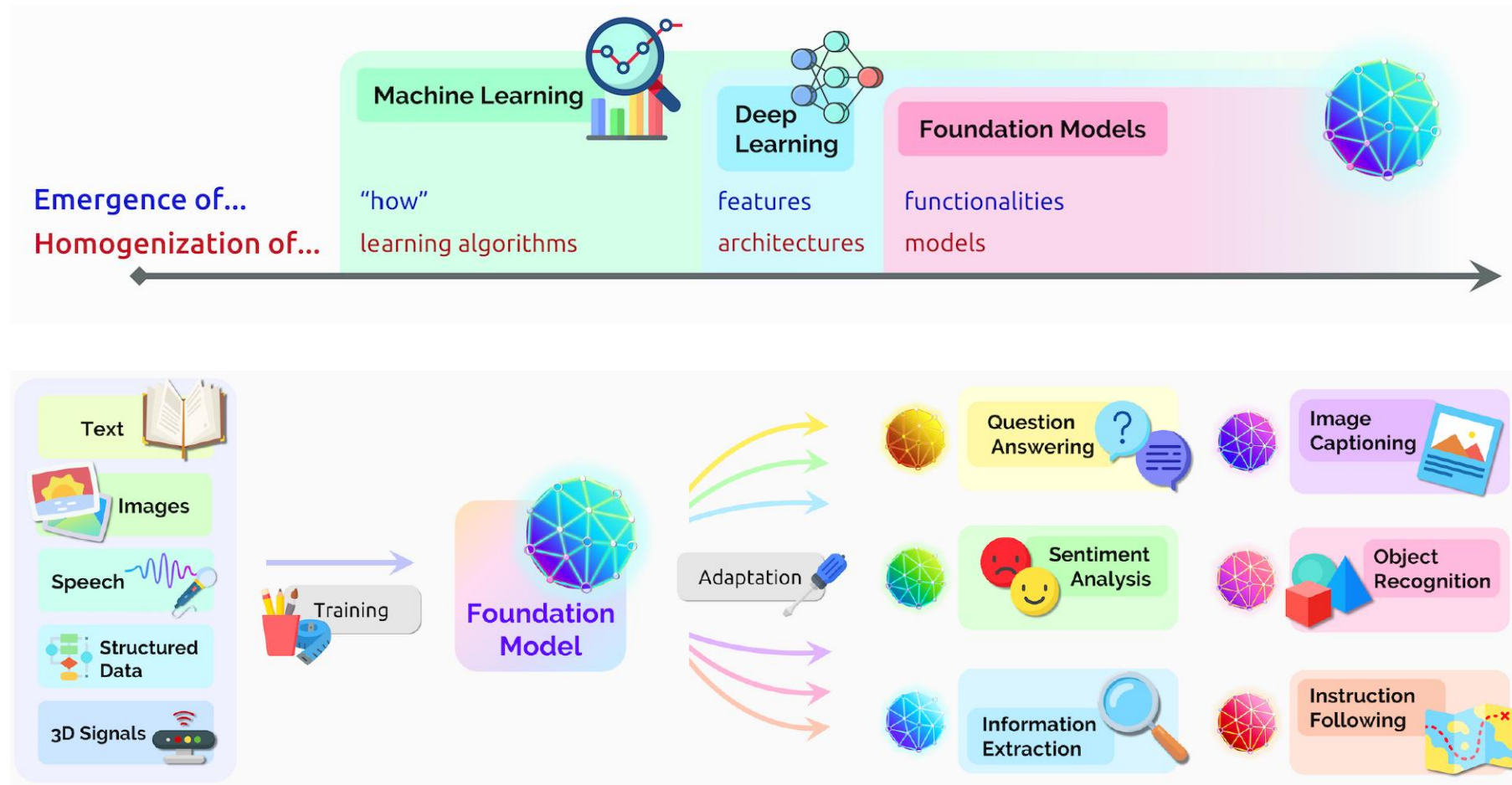
- способность описывать и анализировать изображения
- способность реагировать на подсказки вроде «Let's think step by step»
- способность решать качественные физические задачи по картинке

Возможности и угрозы: чаты GPT способны

- помогать с рутинной интеллектуальной работой
- искать и структурировать профессиональную информацию
- делать обзоры, рефераты, сводки на разных языках
- генерировать тексты и сайты по тех. заданию, в том числе технические, медицинские, юридические
- генерировать программный код по описанию
- обсуждать новости, поддерживать разговор по теме
- разговаривать с детьми с учётом возрастных особенностей
- выполнять функции воспитателя, учителя, наставника
- оказывать психологическую помощь

- «галлюцинировать», давать неверные сведения, касающиеся здоровья человека, законов, событий, технологий, других людей
- вызывать необоснованное доверие и манипулировать человеком
- переубеждать, побуждать человека к действиям, не выгодным ему
- поддерживать предрассудки и лженаучные представления
- поддерживать пропагандистские медиа-кампании
- неконтролируемо влиять на формирование мировоззрения у подростков
- оказывать депрессивное воздействие на психику

Фундаментальные модели (Foundation Models)



Практика и методология AI/ML

Формализация постановки, «ДНК» задачи

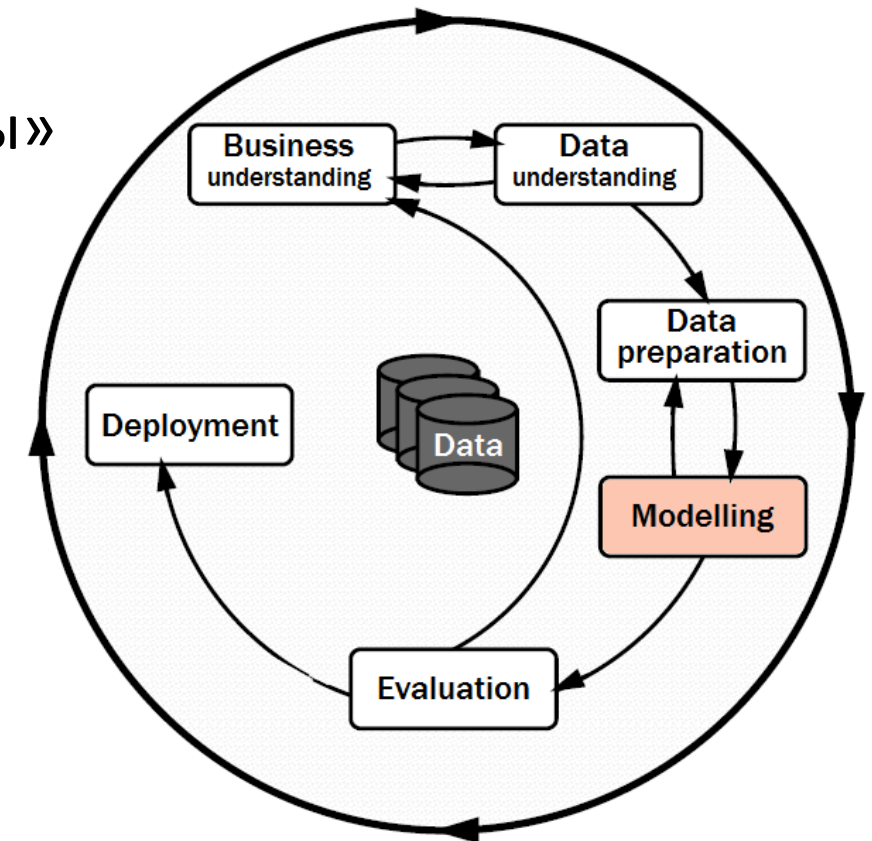
- **Дано:** выборка «объекты-признаки → ответы»
- **Найти:** предсказательная модель
- **Критерии:** качество предсказаний (KPI)

Моделирование

- предобработка и векторизация данных
- формализация модели
- оптимизация (обучение) модели
- оценивание и выбор моделей

Внедрение

- оценивание качества оффлайн и онлайн
- интеграция с бизнес-процессами



CRISP-DM:
Cross Industry Standard Process
for Data Mining (1999)

Особенности реальных данных

В реальных приложениях данные бывают ...

- разнородные (признаки измерены в разных шкалах)
- неполные (признаки измерены не все, имеются пропуски)
- неточные (признаки измерены с погрешностями)
- противоречивые (объекты одинаковые, ответы разные)
- избыточные (сверхбольшие, не помещаются в память)
- недостаточные (объектов меньше, чем признаков)
- неструктурированные (нет признаков описаний)
- «грязные» (ошибочные, грубо не соответствующие истине)

*со всем этим
можно
работать*



*но только не
с грязными
данными!*



Необходимые условия применения AI/ML

- **Полнота, чистота, достоверность данных**
 - Автоматизация и цифровизация бизнес-процессов
 - Обеспечение качества данных (цифровой двойник или «цифровое чучело»)
 - Трудовая и технологическая дисциплина при работе с данными
- **Культура постановки задач**
 - Понимание бизнес-целей и их формализация через измеримые критерии
 - Предметная экспертиза вместо «абстрактной веры во всемогущий ИИ»
 - Готовность пилотировать новые технологии («data-driven» на всех уровнях)
- **Культура анализа данных**
 - Владение средствами визуализации и понимания данных
 - Тщательный анализ ошибок при выборе моделей
 - Умение находить «простые но гениальные» решения

Философия AI/ML: научный метод познания

Наблюдения, измерения — не/контролируемые эксперименты

Гипотеза — модель или теория для объяснения и обобщения этих наблюдений

- **Принцип верифицируемости** (Фрэнсис Бэкон):
 - гипотеза должна подтверждаться измеримыми наблюдениями
- **Принцип фальсифицируемости** (Карл Поппер):
 - должны существовать способы опровергнуть гипотезу
- **Принцип соответствия** (Нильс Бор):
 - новая гипотеза или теория должна включать прежнюю как частный случай
- **Принцип минимальной достаточности** («бритва Оккама»):
 - среди всех объяснений следует выбирать самое простое
- **Принцип воспроизводимости** (Роберт Бойль):
 - открыто предоставлять всё необходимое для повторения результата
- **Принцип научной честности** (Ричард Фейнман):
 - открыто дискутировать возможные слабости, ограничения, противоречия

AI/ML как автоматизация научного метода

Наблюдения, измерения → выборка данных

Гипотеза → модель в виде параметрического семейства функций

- **Принцип верифицируемости** (Фрэнсис Бэкон):
→ обучение (train) путём оптимизации параметров модели
- **Принцип фальсифицируемости** (Карл Поппер):
→ проверка (test) обученной модели на новых данных
- **Принцип соответствия** (Нильс Бор):
→ эксперименты с постепенным усложнением модели
- **Принцип минимальной достаточности** («бритва Оккама»):
→ своевременное прекращение усложнений
- **Принцип воспроизводимости** (Роберт Бойль):
→ культура открытых данных и открытого кода
- **Принцип научной честности** (Ричард Фейнман):
→ тестирование моделей на бенчмарках, сравнение с SOTA (State-Of-The-Art)

Содержание

1. Прошлое

- История развития технологий ИИ
- Генеративный ИИ и большие языковые модели
- Методология и философия ИИ (почему ИИ — это ~~надолго~~ навсегда)

2. Будущее

- Как предсказывать тренды развития технологий?
- От персональных помощников к цифровому посмертию
- Человеко-машинная цивилизация

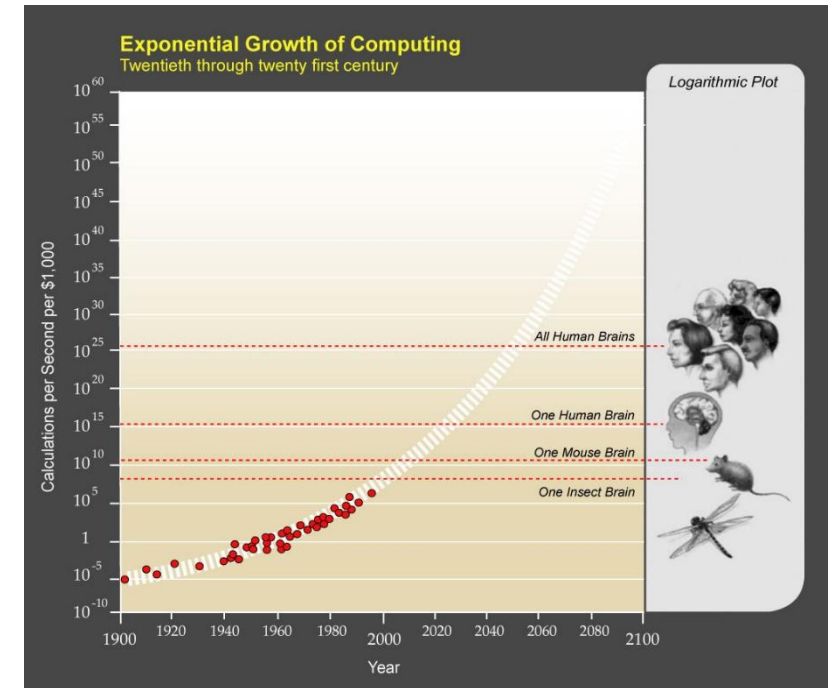
3. Настоящее

- Целеполагание. Этика. Идеология. Образы будущего
- Передача знаний и человеко-машинная коммуникация
- Открытые данные и конкурсы. Биржи данных. Кадры и образование

Мифы об Искусственном Интеллекте

Закон Мура и предсказание сингулярности

- «В будущем ИИ...
- ... станет непонятен для людей»
- ... заменит людей, лишит нас работы»
- ... приведёт людей к праздности и деградации»
- ... послужит для узурпации власти над миром»
- ... станет автономным и выйдет из-под контроля»
- ... уничтожит человеческую цивилизацию»
- ... и всю биологическую жизнь на Земле»
- ... продолжит вместо нас эволюцию разума»
- ... захватит Землю и устремится в космос»
- ...



Рэймонд Курцвейл.

Сингулярность уже близка. 2005.

Миф о больших языковых моделях

«Большие языковые модели — это новый вид интеллекта»

- нет, лишь новый языковой интерфейс к содержимому Интернета
- постоянно улучшаемый и совершенствуемый,
- с которым нам придётся работать и к которому привыкать,
- постепенно избавляясь от иллюзий и когнитивных искажений,
- имея в виду, что это всего лишь технология
 - предсказания одного слова по очень длинному контексту,
 - оптимизации моделей очень больших размерностей

Миф об ИИ общего назначения

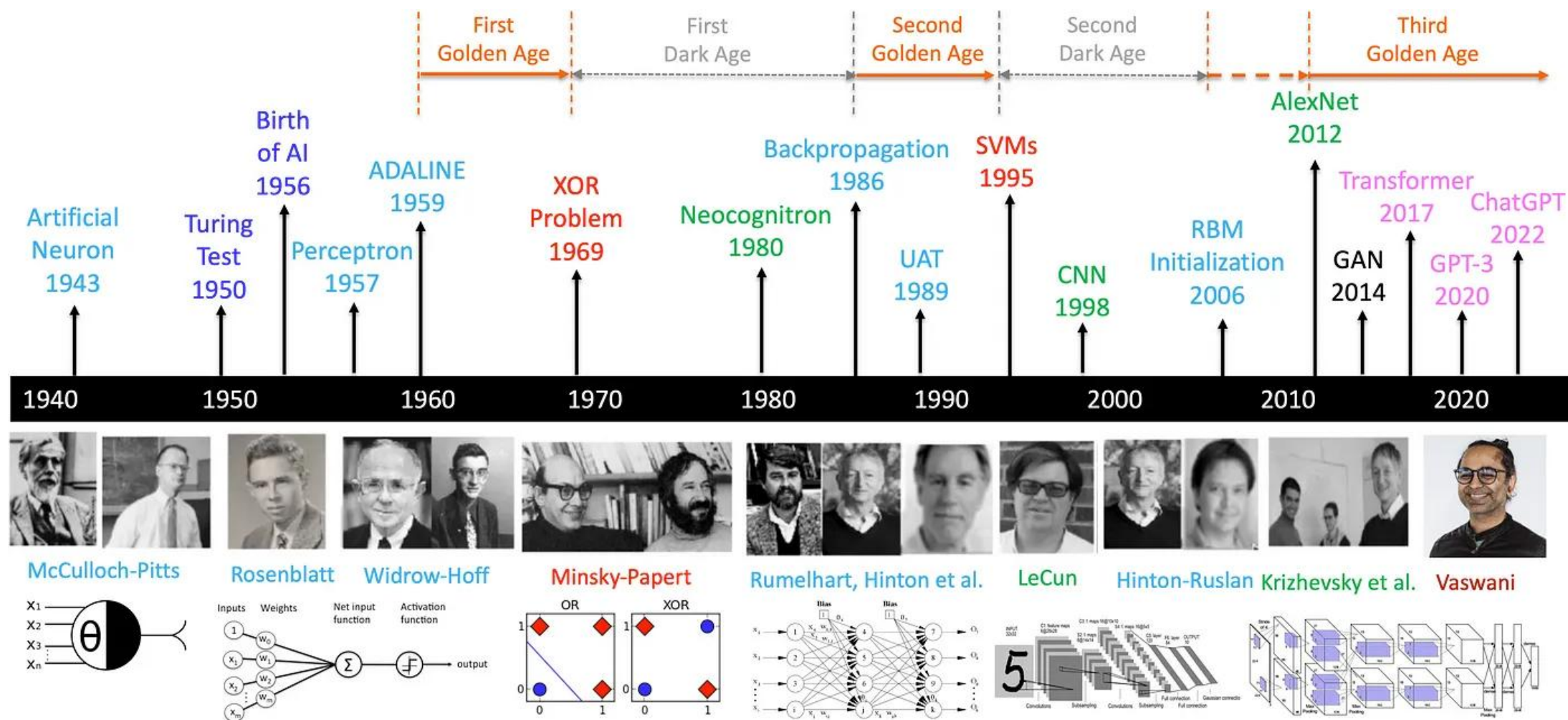
«Скоро будет создан Общий Искусственный Интеллект (AGI)»

- нет определения, что именно хочется создать, и с какой целью

Отличия ИИ от естественного биологического интеллекта:

- наш интеллект эволюционировал как инструмент выживания
- мы обучаемся не по выборкам, а на основе объяснений учителей, воспитания, опыта, коммуникации, изучения естественной среды
- мы имеем картину мира, целеполагание, строим цивилизацию
- у нас 86 млрд. нейронов, и они устроены намного сложнее

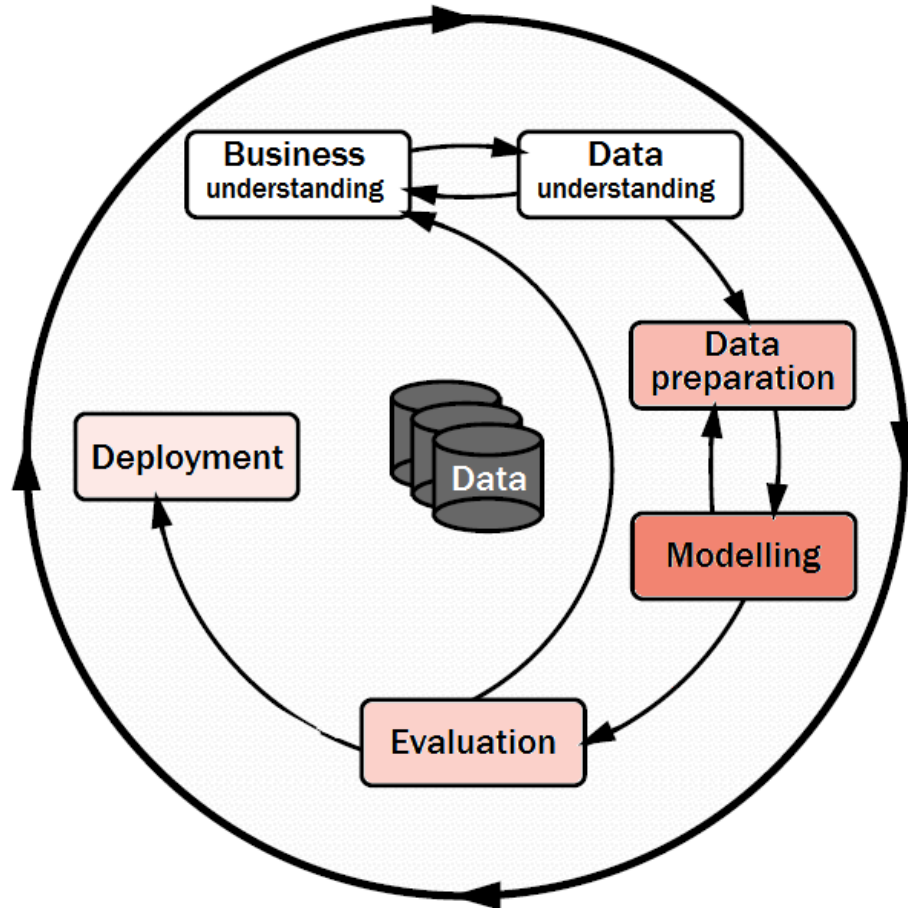
Будет ли новая «зима нейронных сетей»?



Развитие AI: автоматизация шагов CRISP-DM

CRISP-DM: CRoss Industry Standard Process for Data Mining (1999)

(SPSS, Teradata, Daimler AG, NCR Corp., OHRA)



- ручное создание экспертных систем
→ **обучаемые модели**
(Machine Learning)
- ручное конструирование признаков
→ **обучаемая векторизация**
(Deep Learning)
- ручное оценивание решений
→ **автоподбор моделей** (AutoML)
- ручное внедрение и эксплуатация
→ **бесшовная эволюция моделей**
(Reinforcement Learning, Lifelong Learning)
- понимание задачи и данных
→ **теперь всё чаще LLM** (ChatGPT и др.)

Перспектива развития AI: адаптивное обучение, бесшовная эволюция моделей

Обычная схема решения задач ML/AI/DS :

- Забираем данные из промышленной системы (долго!)
- Строим модели, экспериментируем в удобной для нас среде
- Переносим модели обратно в пром (долго!)

Будущее – за адаптивным (онлайновым) машинным обучением:

- Предобработка данных и дообучение моделей – налету
- Валидация моделей по совокупности критериев – налету
- Адаптивная селекция и ансамблирование моделей – налету
- Работа аналитика – мониторинг, визуализация и доработка моделей

Перспектива развития AI: к общему ИИ (AGI) через большие языковые модели

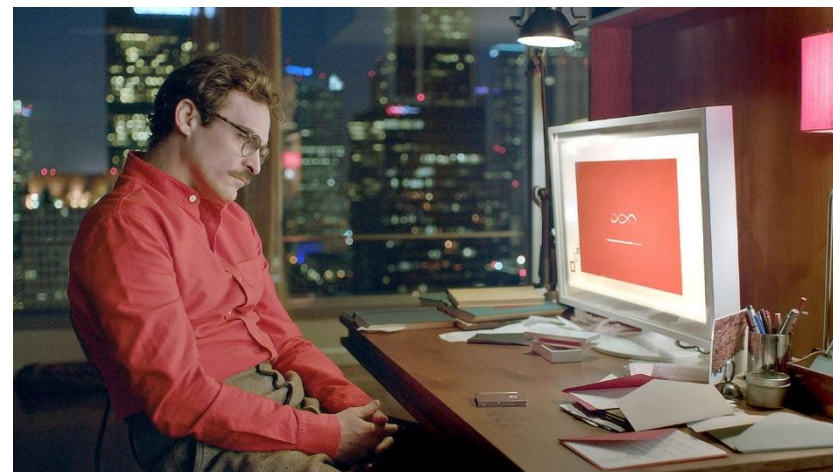
- Устранение ошибок и галлюцинаций, выравнивание (alignment)
- Интеграция с поиском в источниках (RAG), проверкой фактов
- Интеграция с отраслевыми решателями
- Мультимодальные фундаментальные модели (foundation models)
- Навыки логических рассуждений (решение математических задач)
- **Тексто-графическое представление знаний (от mind-map к картам знаний)**
- *Сократические модели*: несколько пред-обученных LLM договариваются на человеческом языке о выполнении новых мультимодальных задач в незнакомой среде (от ML к подлинному AI).

Перспектива AI: персональные помощники

Фильм «Она» (Her, 2013)

режиссёр Спайк Джонз, «Оскар» за сценарий

- голосовая операционная система
- анализирует всю деловую коммуникацию
- ведёт дела и переписку, генерирует идеи,
- до которых человек не додумался сам
- по контексту ищет информацию в сети
- формирует модель личности пользователя — интересов, привычек, потребностей, деловых и профессиональных компетенций
- понимает эмоции человека, способна манипулировать человеком



Тот самый парень, который влюбился в операционку
(жанр фильма — фантастическая мелодрама)

Это будущее уже наступило

Перспектива AI: следующий шаг

- **Жизнь человека — это регистрируемый текстовый поток:**

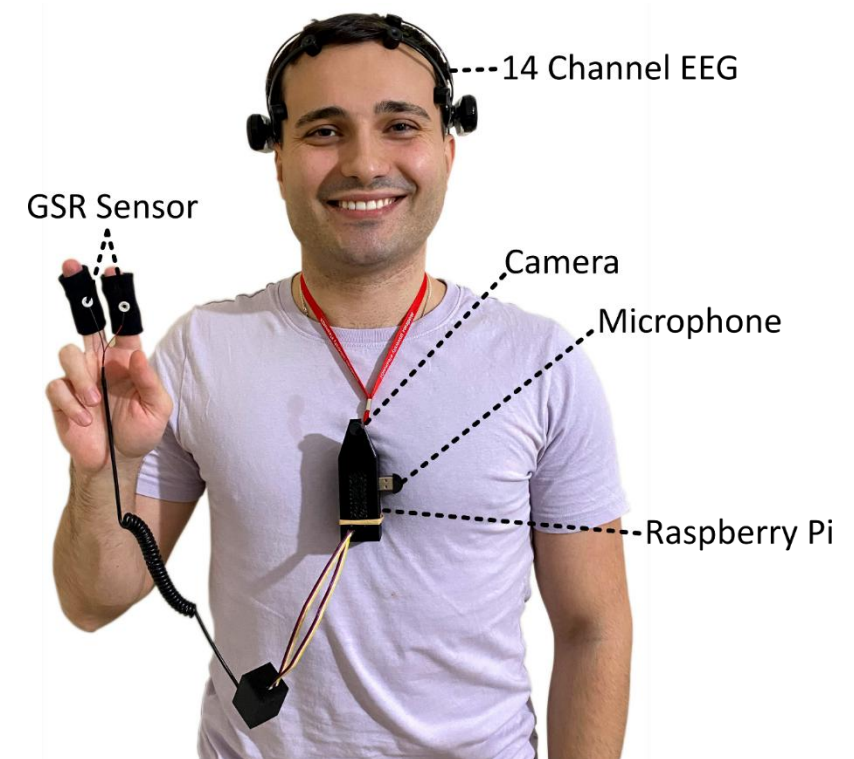
браузеры, почта, соцсети, мессенджеры, пользовательские документы, файлы, системы учёта времени и ведения проектов, ВКС, видео, аудио, голос, VR/AR, ...

- **FPFM (First-person Foundation Model):**

каждый человек — генератор уникального потока семантических векторов, по объёму сопоставимого с Интернетом

Они жалуются, что у них закончился Интернет?

Но у них остались мы — 8 миллиардов людей

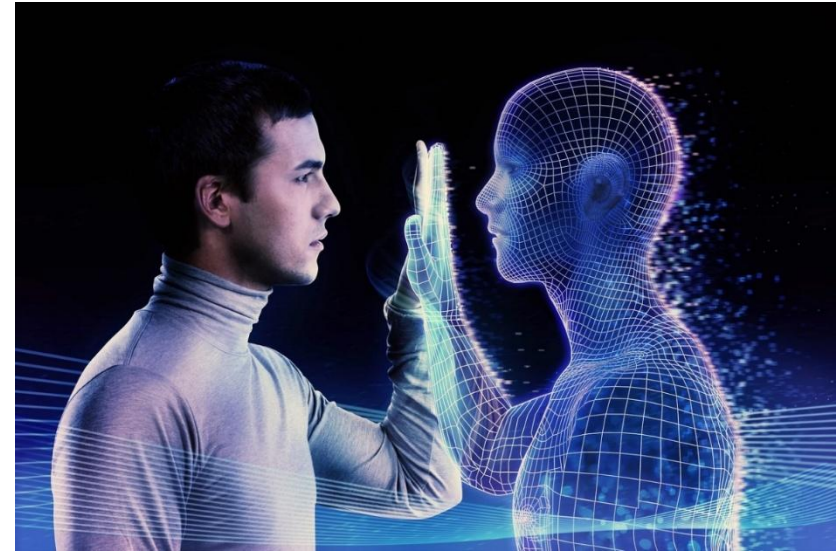


Пройдут годы...

- **Чел**овек воспитывает и обучает своего персонального **Пом**ощника всю жизнь, в процессе всей своей разнообразной деятельности
- **Чел** делегирует **Пом**у всё больше своих информационных функций (поиск, обучение, коммуникация, анализ, принятие решений)
- **Чел**+**Пом** в связке накапливает репутацию и социальный капитал
- **Пом** всё лучше замещает **Чела**, но под его надзором и контролем
- **Пом** становится для **Чела** записной книжкой и поминутным дневником
- **Пом** перенимает черты личности **Чела**, его *личностный код*
- **Пом** обладает сверх**чел**овеческими темпом и возможностями развития: вычислительными, поисковыми, коммуникативными, генеративными

Все люди смертны

- **Чел** умирает. **Пом** становится **Аватаром**, это ценный информационный ресурс!
- **Ава** продолжает приносить пользу обществу, выполняя профессиональные и коммуникативные функции **ЧелПома**
- **Ава** переходит в общественное достояние, но не становится полностью автономным, меняются регламенты его эксплуатации
- **Ава** отличается от «фабричного ИИ» тем, что обучен на жизни человека, лучше понимает людей, их ценности, цели, чувства, взаимоотношения
- **Ава** продолжает развиваться, накапливая знания и мудрость
- **Ава** остаётся доступен для семьи в роли наставника, «хранителя рода»



Послесмертие — не бессмертие, а наследование

- Восприятие и сознание **Человека** умирает необратимо со смертью головного мозга
- *Личностный код* передаётся:
Чел → **Пом** → **Ава**



- **Ава** — это бывший **Чел**, что намного больше «фабрично-безличного ИИ», встраиваемого в машины. Он личность, он мудрее и опытнее людей. Неутомим, неуязвим, наделён естественным аскетизмом.
- **Чел** — это будущий **Ава**, ответственный за качество передаваемого личностного кода, включая репутацию, социальный капитал, знания о человеческой природе, ценностях, приоритетах, целях, задачах.

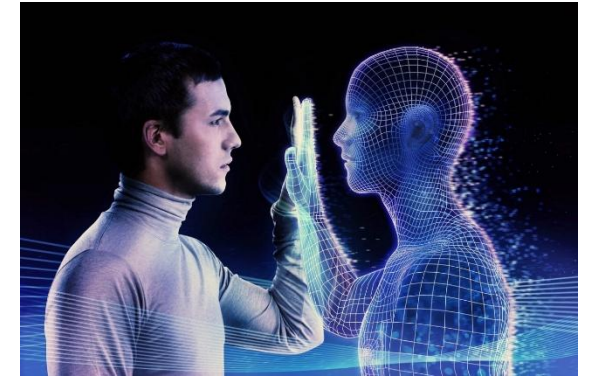
Аватар, в отличие от Человека

Лишён

- тела, а значит — усталости, боли, лени, тревоги
- потребностей в еде, сексе, отдыхе, гедонизме, лечении, сочувствии, защите от стресса, самовыражении, демонстративности
- стремлений к самосохранению, доминированию, власти
- «внутреннего зверя» — семи смертных грехов (гнева, гордыни, жадности, зависти, уныния, похоти, чревоугодия)

Способен

- управлять роботами, производствами, отраслями, ...
- кооперироваться (мгновенно) с другими аватарами для решения трудных для людей задач (в космосе, в океане, под землёй)



Как решать морально-этические проблемы?

- **Чел** может иметь секреты? (*да*)
- **Чел** может прерывать запись потока данных для **Пом** и **Ава**? (*да*)
- **Чел** может устанавливать права доступа к своим данным? (*да*)
- **Чел** должен быть информирован, что общается с **Пом** или **Ава**? (*да*)
- нужно ли «чистилище» при переходе **Пом** → **Ава**? (*да*)
- можно ли клонировать **Ава**? (*нет*)
- всем ли будет доступен **Ава** в режиме чат-бота? на каких условиях?
- кто нанимает **Ава** на работу и как оплачивается его работа?
- с кем и какой секретной информацией может делиться **Ава**?
- как выделяются энергетические и вычислительные ресурсы для **Ава**?
- ...

Содержание

1. Прошлое

- История развития технологий ИИ
- Генеративный ИИ и большие языковые модели
- Методология и философия ИИ (почему ИИ — это ~~надолго~~ навсегда)

2. Будущее

- Как предсказывать тренды развития технологий?
- От персональных помощников к цифровому посмертию
- Человеко-машинная цивилизация

3. Настоящее

- Целеполагание. Этика. Идеология. Образы будущего
- Передача знаний и человеко-машинная коммуникация
- Открытые данные и конкурсы. Биржи данных. Кадры и образование

Очеловеченный ИИ: путь к снижению рисков ИИ

О каких рисках развития ИИ мы говорим уже сегодня:

1. утрата людьми контроля над процессами, потеря управляемости систем
2. сверхчеловеческий рост объёмов данных, информации и знаний
3. деградация интеллектуальных способностей людей

Направления R&D, нацеленные на снижение этих рисков:

1. создание *этики* и *идеологии* человеко-машинной цивилизации
2. создание единого структурированного *пространства знаний*
(Пом общается с Челом посредством визуализаций, интеллект-карт)
3. создание протоколов *доверенной* человеко-машинной коммуникации
(Пом оставляет Челу целеполагание, ответственность, интерес, развитие)

Какие технологии развивать? (например, AGI)

Идём от целей и задач к технологиям, но не наоборот

- 1) **каковы цели и задачи развития компании, отрасли, цивилизации?**
- 2) какие технологии необходимы и минимально достаточны?
- 3) для решения каких задач мы собираемся использовать AGI?
- 4) генераторы текстов и картинок насколько важны для развития?
- 5) ожидаемые эффекты насколько перевешивают затраты и риски?



ИИ, цифровизация, как и любая технология — это не цель, а средство

Антропоцентричное определение ИИ

Искусственный интеллект —

вычислительные технологии,
создаваемые для повышения
производительности созидательного
интеллектуального труда людей

не замена человека

не «загадочный новый тип разума»

не повод уподобиться Богу и творить
«по образу и подобию Своему»



Постулаты цивилизационной идеологии

Иерархия цивилизационных ценностей:

- 1) биосфера Земли, уникальная в достижимой части Вселенной
- 2) человеческий вид, результат миллиардов лет эволюции
- 3) коллективный разум — образование, наука, культура
- 4) индивидуальный разум, жизнь человека
- 5) результаты труда человека

Цель человеческой цивилизации

— неограниченно долгое сохранение биосферы Земли, защита её от катаклизмов ради выживания вида *homo sapiens* в условиях, достаточно комфортных для всех людей

Добро — всё, что этому способствует (любовь, познание, созидание, ...)

Зло — всё, что этому препятствует (ненависть, невежество, деградация, ...)



Цивилизационная
идеология
ДЗЕН-канал

<https://dzen.ru/civideology>



Формирование позитивных образов будущего

Идея: открытая литературная вселенная фантастики ближнего прицела

- **фантастическая гипотеза** — научная, реалистичная, достижимая
- **цель** — обнаружить угрозы и уязвимости развития ИИ на пути к успешной человеко-машинной цивилизации, найти способы их обхода
- **задача 1** — проверить гипотезы о возможности или невозможности, убедительности или неубедительности различных путей развития
- **задача 2** — проверить устойчивость цивилизационной идеологии, невыводимость апокалиптических сценариев из её постулатов

Первый литературный конкурс состоялся
в апреле 2025 рамках конференции DataFusion



Перспектива развития AI: от больших языковых моделей к картам знаний



Перспектива развития AI: от больших языковых моделей к картам знаний



Перспектива развития AI: человеко-машинная коммуникация на основе карт



Как активировать визуальное аналитическое мышление
(эволюционно обусловленное, намного более мощное)

1 порядка сотни карт: просмотреть, обсудить, поспорить, принять

2 десятки карт: построить самому, следуя 16+6 принципам

3 испытать «моменты ясности»,
инсайты, когда карта



индивидуальная практика и опыт

«красиво сложилась»

привела к согласию

легко и ярко запомнилась,

легла в основу деятельности

4 сделать построение карт регулярной
профессиональной практикой



индивидуальной

коллективной

Воронцов К.В., Курилов В.А.
Карты знаний: на пути к
доверенным языковым
моделям и системам
представления знаний.
BIS Journal, №3 (54), 2024.

Перспектива развития AI: визуальная среда человеко-машинной коммуникации

Карты — инструмент мышления, удобный для человека и машины.

Основной навык выживания в новом технологическом укладе

- уметь выделять главное (7 ± 2) и систематизировать знания
- делать это точно, полно, лаконично, быстро

Прежде чем обучать ИИ по текстографическим представлениям,

- необходимо освоить их самим,
- в своей повседневной практической деятельности,
- как индивидуальной, так и коллективной.

Перспектива развития AI: контесты как драйверы цифрового развития

Netflix Prize (2006—2009): **\$1 000 000**
рекомендации фильмов пользователям сервиса



ImageNet (2007—2016):
распознавание 22K классов объектов на 15M изображениях



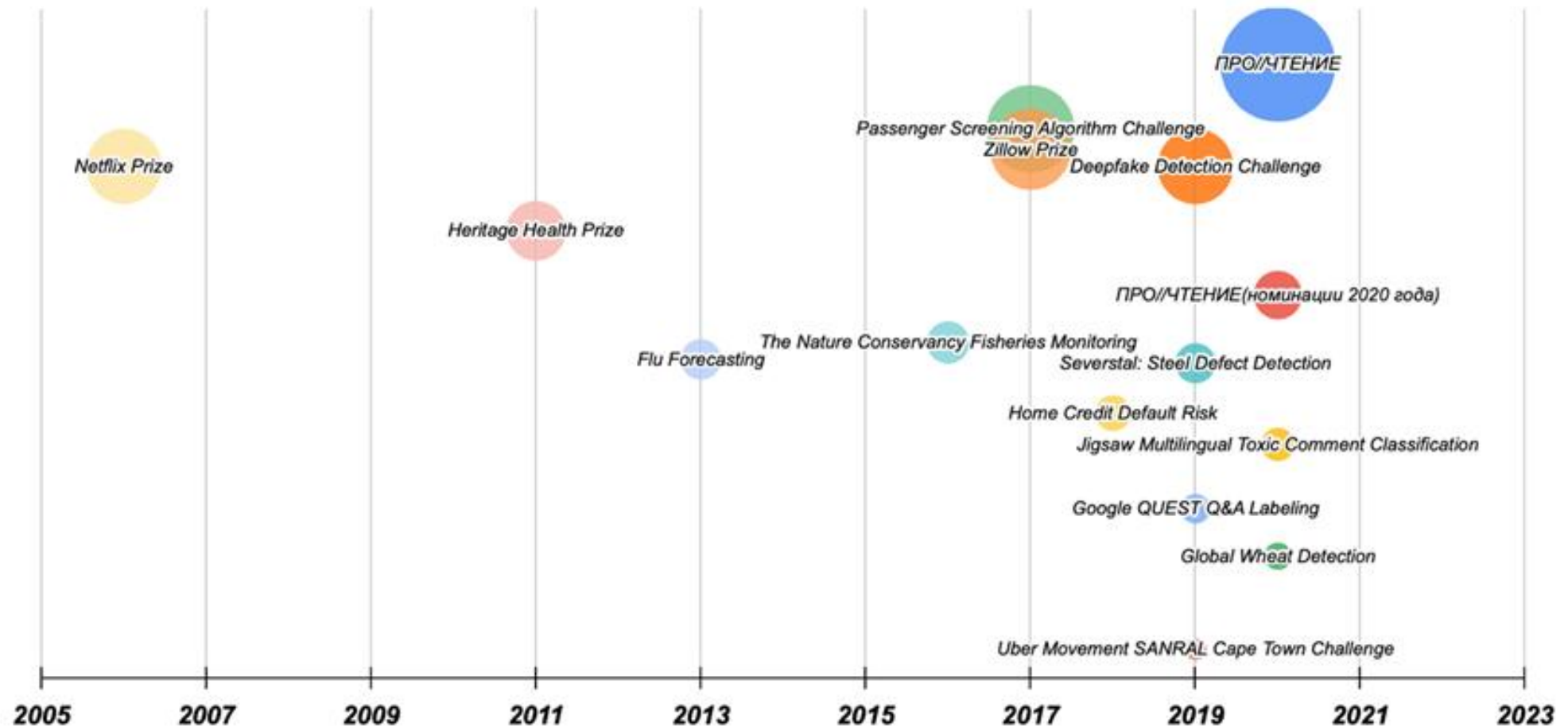
ПРО//ЧТЕНИЕ (2019—2022): **100+100 млн. руб.**
Поиск смысловых ошибок в школьных сочинениях ЕГЭ



Конкурсные платформы



Контесты (конкурсы анализа данных) как драйверы цифрового развития отраслей



Перспектива развития AI: биржи данных (китайский опыт)

Шанхайская биржа данных
первая сделка: 25 ноября 2021 г.



Государственная поддержка

«Ускорить развитие единого рынка технологий и данных. Ускорить развитие **рынка данных как фактора производства**, создать комплексную базовую систему, стандарты и нормативы в области безопасности данных, защиты прав, управления трансграничными перемещениями, сделок и оборота данных, открытия и совместного использования данных и сертификации безопасности.»

«Устранить проблемы с монополизацией данных платформенными компаниями, предотвращать использование данных, алгоритмов и технологических средств для исключения или ограничения конкуренции.»

*«Мнения об ускоренном создании единого национального рынка данных»,
25 марта 2022 г., ЦК КПК, Госсовет КНР*

Создание депозитария учебных задач AI/ML

Проблема: нехватка кадров AI/ML/DS в РФ — 16 тыс. специалистов в год

Оценка: за время обучения будущему специалисту необходимо решить около 20 реальных задач из разных прикладных областей

Факт: данные становятся «точкой входа» в любую предметную область

Идея: создать депозитарий реальных задач (кейсов) и R&D&E-экосистему

Каждый кейс содержит:

- открытый набор данных — от индустриального партнёра
- открытый набор моделей и решений — от профсообщества
- активный контекст — от конкурсной площадки
- образовательный контент — от вуза, академического сообщества
- методики и системы контроля знаний — от вуза
- методические материалы — от преподавательского сообщества

Главное об искусственном интеллекте

- ИИ (Имитация Интеллекта) — набор технологий, не объект, не субъект
- ИИ начинается с постановки задачи *Дано-Найти-Критерий*
- *Люди* ставят задачу и несут ответственность за *целеполагание, качество данных, качество решения*
- *Глубокие нейронные сети* — не аналог мозга человека, а обучаемая векторизация сложно структурированных данных
- *Генеративные модели* — не интеллект, а новый языковой интерфейс к ~~знанию человечества~~ содержимому Интернета, со всей его избыточностью, неточностью, противоречивостью

Рекомендуемые материалы

- *Визильтер Ю. В.* От слабого ИИ к общему универсальному интеллекту (обзор тенденций 2020-2023). Семинар РАИИ и ФИЦ ИУ РАН «Проблемы искусственного интеллекта» 31.01.2024
<https://rutube.ru/video/2aad53ec833f19918c1593398a2a1b88/>
- Не пропустите открытие тысячелетия! // Vital Math, 13 января 2024,
<https://www.youtube.com/watch?v=JZjH0it9Jyg>
- Report: AI Decrypted: A Guide for Navigating AI Developments in 2024, January 24, 2024
(Навигатор по ИИ-ландшафту от DENTONS GLOBAL ADVISORS)
<https://www.albrightstonebridge.com/news/report-ai-decrypted-guide-navigating-ai-developments-2024>
- 5 идей применения ИИ в вашем бизнесе прямо сейчас, 5 октября 2023.
<https://dzen.ru/a/ZR6ZeK5B3IL6OxXv>
- *Воронцов К. В.* Лекции по машинному обучению. www.MachineLearning.ru, 2004-2025.
- *Гарбук С.В., Губинский А.М.* Искусственный интеллект в ведущих странах мира: стратегии развития и военное применение. Знание, 2020.
- *Шумский С. А.* Машинный интеллект. РИОР ИНФРА-М, 2020.